IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 9, Issue 3, March 2020

Oral Cancer Detection and Level Classification Through Machine Learning

Jyoti Rathod[1], Shraddha Sherkay[2], Harshal Bondre[3], Rohit Sonewane[4], Devika Deshmukh[5]

Student, Department of Computer Technology, Yeshwantrao Chavan College of Engineering, Nagpur, India[1,2,3,4]

Professor, Department of Computer Technology, Yeshwantrao Chavan College of Engineering, Nagpur, India[5]

Abstract: Oral cancer is the most common type of cancer. It was an irrecoverable disease but now the progress in technology has made it curable if it is diagnosed in early stages. Oral cancer is increase in the number of cells which has the capability to affect its neighbor cells or tissues. It happens when cells divide out of control and form a growth, or tumor. In spite of having various advancements in fields like radiation therapy and chemotherapy the mortality rate is persistent. Therefore, early detection of cancer is important. In this paper we are using Machine Learning as domain for detection of oral cancer considering the datasets of a victim. Then it is classified using apriori algorithm. We are developing a health sector application which also uses Data Mining and data extraction for prediction techniques, classification rules for oral cancer prediction and uses association rules to perceive the relationship between the oral cancer attributes.

Keywords: Oral cancer detection, machine learning, web-crawler,data mining technique, association rule mining, apriori algorithm.

I INTRODUCTION

Cancer has been characterized as a heterogeneous disease consisting of various subtypes. Oral cancer is the most dangerous type of cancer. It is caused in various parts such as lips, tongue, hard and soft palate and floor of mouth. Oral cancer is caused due to tobacco use of any kind, including cigars, pipes, chewing tobacco, snuffs, or alcohol consumption. Its detection and diagnosis is very important or else it can be fatal. Therefore early diagnosis of a cancer type have become a necessity in cancer research. The ability of ML tools to detect key features from composite datasets reveals their significance. The predictive models discussed here are based on various supervised ML techniques as well as on different input features and data samples[3]. Given the rapidly growing trend on the application of ML methods in cancer research, we present here the most recent publications that employ these techniques as an aim to model oral cancer image classification or patient outcomes.

II LITERATURE REVIEW

Previously cancer was an incurable disease, but now with the advancement in technology it has been successful in becoming a curable disease. Oral cancer is the unstoppable increase in the number of cells or mutation that is formed and has the capability to affect the neighbouring tissues. In this paper different algorithms of data mining will be used to detect oral cancer.

From paper [1], they proposed approach WEKA is applied with ten cross validations to calculate and collate output. WEKA consists of a large variety of data mining machine learning algorithms. First we have classified the oral cancer dataset and then analyzed various data mining methods in WEKA through Explorer and Experiment interfaces. we get oral cancer detection about data mining tool.
From paper [2] we get details about Computer Aided Diagnostics of Facial & Oral Cancer. Cancer is the leading cause of death. It's important to find oral cancer early when it can be treated more successful. Content-based image retrieval (CBIR) is a promising method for computer-aided diagnostics leading early diagnosis. In this paper, we perform FOCT (Facial and Oral Cancer Tracker), our new platform which assist surgeons in decisions regarding new cases by supplying visually similar past cases.
Articles from Computational and Structural Biotechnology Journal are provided about Machine learning applications in cancer prognosis and prediction. Cancer has been characterized as a heterogeneous disease consisting of many different subtypes. The early diagnosis and prognosis of a cancer type have become a necessity in cancer research, as it can facilitate the subsequent clinical management of patients. The importance of classifying cancer patients into high or low risk groups has led many research teams, from the biomedical and the bioinformatics field, to study the application of machine learning (ML) methods [3].

DOI 10.17148/IJARCCE.2020.9335

From paper [4], The main aim of their work is to assess the clinical features, diagnostic procedures and treatment required for oral cancer patients. The staging of the cancer is generally divided into two stages namely, clinical and pathological. In TNM (Tumour, Node, Metasis), a lot of novel prognostic tools have been traced and new methodologies for the prognostic factors have been drastically improved and developed. This paper compares the classification accuracy of the TNM staging system with the aid of Multi-Layer Perceptron (MLP) and Gaussian Mixture Model (GMM) classifiers.
In this article [5] they formulate a Decision Support System (DSS) which integrates a multitude of heterogeneous data (clinical, imaging and genomic), thus, framing all manifestations of the disease. Our primary aim is to identify the factors that dictate Oral squamous cell carcinoma (OSCC) progression and subsequently predict potential relapses (local or metastatic) of the disease. The discrimination potential of each source of data is initially explored separately, and afterwards the individual predictions are combined to yield a consensus decision achieving complete discrimination between patients with and without a disease relapse.
Medical imaging technique, computer-aided diagnosis and detection can make potential changes in cancer treatment. In this research [6] work, they have developed a deep learning algorithm for automated, computer-aided oral cancer detecting system by investigating patient hyperspectral images. METHODS: To validate the proposed regression-based partitioned deep learning algorithm, they compare the performance with other techniques by its classification accuracy, specificity, and sensitivity. For the accurate medical image classification objective, they demonstrate a new structure of partitioned deep Convolution Neural Network (CNN) with two partitioned layers for labelling and classify by labelling region of interest in multidimensional hyperspectral image.
The purpose of retrieving relevant images with appropriate keyword(s) an image crawler [7] is designed and implemented. Here, keyword(s) are submitted as query and with the help of sender engine, images are downloaded along with metadata like URL, filename, file size, file access date and time etc. Later, with the help of URL, images already present in repository and newly downloaded are compared for uniqueness. Only unique URLs are in turn considered and stored in repository. The images in the repository are used to build novel Content Based Image Retrieval (CBIR) system in future. This repository may be used for various purposes. This image crawler tool is useful in building image datasets which can be used by any CBIR system for training and testing purposes.
The analysis of the error backpropagation algorithm, they propose an innovative training criterion of depth neural network [8] for maximum interval minimum classification error. At the same time, the cross entropy and M3CE are analyzed and combined to obtain better results. Finally, we tested our proposed M3 CE-CEc on two deep learning standard databases, MNIST and CIFAR-10. The experimental results show that M3 CE can enhance the cross-entropy, and it is an effective supplement to the cross-entropy criterion. M3 CE-CEc has obtained good results in both databases.

III WORK DONE

Overview

This project is divided into following different modules:

1 Web Scraping
2 Data Training
3 Prediction

1. Web Crawler

Web Scraping (also termed Screen Scraping, Web Data Extraction, Web Harvesting etc.) is a technique employed to extract large amounts of data from websites whereby the data is extracted and saved to a local file in your computer or to a database in table (spreadsheet) format.

Data displayed by most websites can only be viewed using a web browser. They do not offer the functionality to save a copy of this data for personal use. The only option then is to manually copy and paste the data - a very tedious job which can take many hours or sometimes days to complete. Web Scraping is the technique of automating this process, so that instead of manually copying the data from websites, the Web Scraping software will perform the same task within a fraction of the time.

Before we jump into building the model, we need to download all the required oral cancer image dataset by using web scraping, store into both the folders named “cancerous” and “non-cancerous” into our working directory, it may take a while as there are thousands of images in both folders, which is the training data as well as the test dataset.

To extract data using web scraping with python, need to follow these basic steps:

i) Find the URL that you want to scrape.
ii) Inspecting the Page.

DOI 10.17148/IJARCCE.2020.9335

iii) Find the data you want to extract.
iv) Write the code.
v) Run the code and extract the data.
vi) Store the data in the required format.

Fig: Images Download by Using Web Crawler

So, we are using the following libraries for web scraping:

- Beautiful soup: Beautiful Soup is a Python library for pulling data out of HTML and XML files. It works with your favorite parser to provide idiomatic ways of navigating, searching, and modifying the parse tree. It commonly saves programmers hours or days of work.
- Requests: Requests is a Python module that you can use to send all kinds of HTTP requests. It is an easy-to-use library with a lot of features ranging from passing parameters in URLs to sending custom headers and SSL Verification.
- Lxml: lxml is a Python library which allows for easy handling of XML and HTML files, and can also be used for web scraping.

2. Data Training

Image Classification using Convolutional Neural Network — Deep Learning in python. We are building a convolutional neural network that has been trained on few thousand images of cancerous and non-cancerous patients, and later be able to predict if the given image is of cancerous and non-cancerous patients. In this project we are solving an image classification problem, where our goal will be to tell which class the input image belongs to. The way we are going to achieve it is by training an artificial neural network on few thousand images of cancerous and non-cancerous patients and make the NN (Neural Network) learn to predict which class the image belongs to, next it sees an image having cancerous and non-cancerous patients in it.

In training part, keras and numpy libraries are used by running them on tensorflow in backend. Keras is TensorFlow's high-level API for building and training deep learning models. It's used for fast prototyping, state-of-the-art research, and production. Numpy is a library for the Python programming language, adding support for large, multi-dimensional arrays and matrices, along with a large collection of high-level mathematical functions to operate on these arrays. Convolutional neural network is a class of deep neural networks, most commonly applied to analyzing visual imagery. Images are classified based on the output values of Activation function whose output comes out to be 0 or 1. 0 for Cancerous and 1 for Non-Cancerous.

Fig: Cancerous and Non-Cancerous Patient’s Images

3. Prediction

In Prediction part also, keras and numpy libraries are used. Keras is TensorFlow's high-level API for building and training deep learning models. It's used for fast prototyping, state-of-the-art research, and production. In training process, two folders are created i.e. cancerous and non-cancerous and then after that we create another folder module.h5 in which we store images called as weights. After all this stuff we will create another folder and save it named “Oral Cancer Detection”. This “Oral Cancer Detection” folder is being to be connected with module.h5 folder. Then we store the images into an array. Find its X and Y coordinates. In this module mathematical formula are used. Then we are using predict () method on our classifier object to get the prediction. As the prediction will be in a binary form, we will be receiving either a 0 or 1, which will represent a cancerous and non-cancerous patient respectively.

IV FLOW DIAGRAM

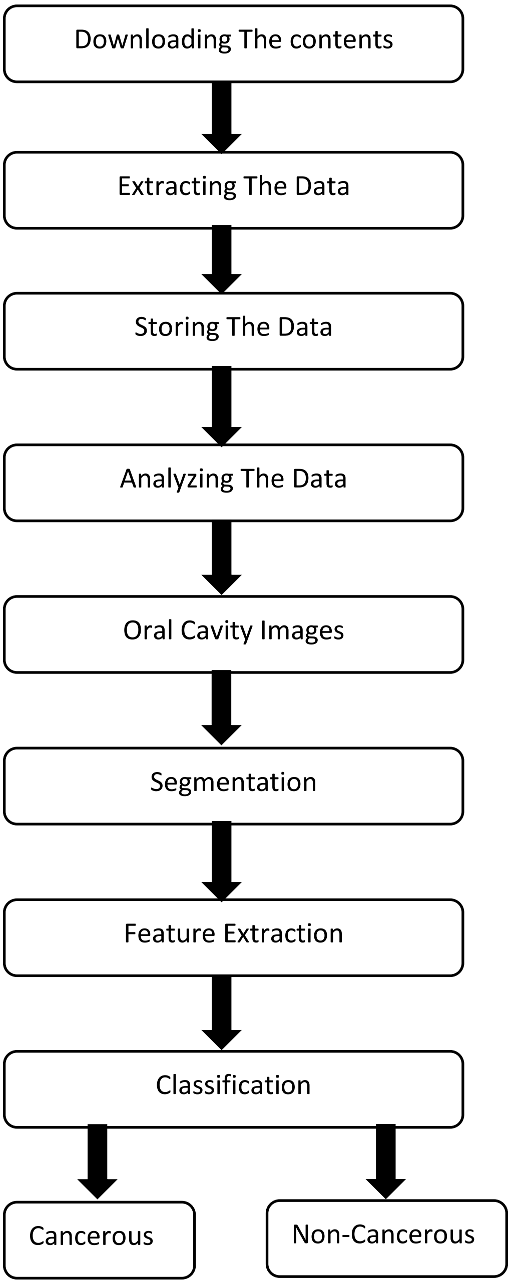


Fig: Flow chart of the system

DOI 10.17148/IJARCCE.2020.9335

V OUTPUT

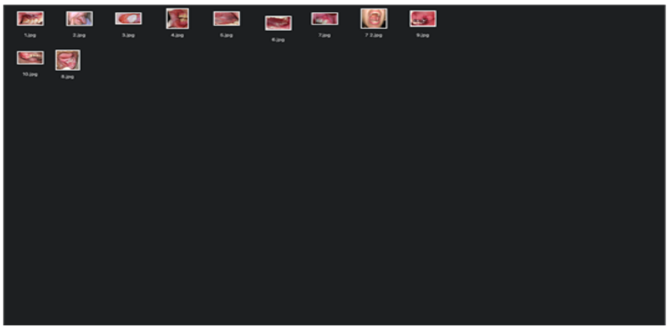
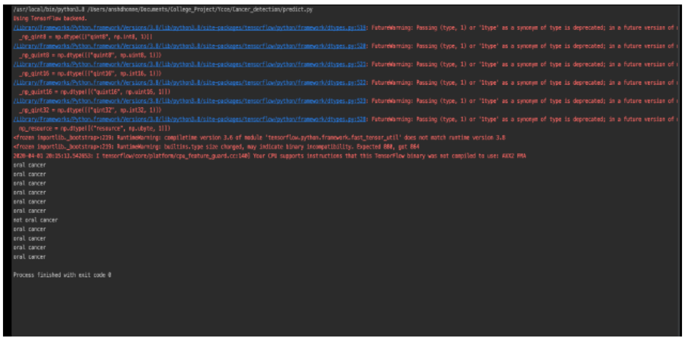


Fig: Test Images Which Is Cancerous or Non-Cancerous



```
/usr/local/bin/python3.8 /Users/anshdhomne/Documents/College_Project/Ycce/Cancer_detection/predict.py
Using TensorFlow backend.
/Library/Frameworks/Python.framework/Versions/3.8/lib/python3.8/site-packages/tensorflow/python/framework/dtypes.py:519: FutureWarning: Passing (type, 1) or '1type' as a synonym of type is deprecated; in a future version of
  _np_qint8 = np.dtype([("qint8", np.int8, 1)])
/Library/Frameworks/Python.framework/Versions/3.8/lib/python3.8/site-packages/tensorflow/python/framework/dtypes.py:520: FutureWarning: Passing (type, 1) or '1type' as a synonym of type is deprecated; in a future version of
  _np_quint8 = np.dtype([("quint8", np.uint8, 1)])
/Library/Frameworks/Python.framework/Versions/3.8/lib/python3.8/site-packages/tensorflow/python/framework/dtypes.py:521: FutureWarning: Passing (type, 1) or '1type' as a synonym of type is deprecated; in a future version of
  _np_qint16 = np.dtype([("qint16", np.int16, 1)])
/Library/Frameworks/Python.framework/Versions/3.8/lib/python3.8/site-packages/tensorflow/python/framework/dtypes.py:522: FutureWarning: Passing (type, 1) or '1type' as a synonym of type is deprecated; in a future version of
  _np_quint16 = np.dtype([("quint16", np.uint16, 1)])
/Library/Frameworks/Python.framework/Versions/3.8/lib/python3.8/site-packages/tensorflow/python/framework/dtypes.py:523: FutureWarning: Passing (type, 1) or '1type' as a synonym of type is deprecated; in a future version of
  _np_qint32 = np.dtype([("qint32", np.int32, 1)])
/Library/Frameworks/Python.framework/Versions/3.8/lib/python3.8/site-packages/tensorflow/python/framework/dtypes.py:528: FutureWarning: Passing (type, 1) or '1type' as a synonym of type is deprecated; in a future version of
  np_resource = np.dtype([("resource", np.ubyte, 1)])
<frozen importlib._bootstrap>:219: RuntimeWarning: compiletime version 3.6 of module 'tensorflow.python.framework.fast_tensor_util' does not match runtime version 3.8
<frozen importlib._bootstrap>:219: RuntimeWarning: builtins.type size changed, may indicate binary incompatibility. Expected 880, got 864
2020-04-01 20:15:13.542653: I tensorflow/core/platform/cpu_feature_guard.cc:140] Your CPU supports instructions that this TensorFlow binary was not compiled to use: AVX2 FMA
oral cancer
oral cancer
oral cancer
oral cancer
oral cancer
oral cancer
not oral cancer
oral cancer
oral cancer
oral cancer
oral cancer

Process finished with exit code 0
```

Fig: Output of The Test Images Cancerous or Non-Cancerous

VI RESULT

The result of this project is to detect the cancerous cells in the oral cavity and to classify the cancer affected position to give results for easier approach to the doctors to start their treatment efficiently. The project is applying various techniques in the field of computer technology using the web-crawler and deep learning convolution neural network is to classify the cancerous and non- cancerous patient's images are spreading to the other parts of the system.

VII CONCLUSION

In diagnosis of Oral Cancer, the staging is one of the important tasks to be performed by medical practitioners of the cancer field. Hence it is important to classify different stages in Oral Cancer to give effective treatment for the cancer patient. We compared the obtained results from another traditional medical image classification algorithm. From the obtained result, we identify that the quality of diagnosis is increased by proposed regression-based partitioned CNN learning algorithm for a complex medical image of oral cancer diagnosis.

VIII ACKNOWLEDGMENT

Presentation, inspiration and motivation have always played a key role in success of any venture. We pay our deep sense of gratitude to **Prof. Gauri Dhopavkar (HOD)** of Computer Technology Department, Yeshwantrao Chavan College of Engineering to encourage us to the highest peak and to provide us the opportunity to prepare the project. We are immensely obliged to our teachers for their elevating inspiration, encouraging guidance and kind supervision in completion of our project. We feel to acknowledge our indebtedness and deep sense of gratitude to our guide **Prof. Devika Deshmukh** whose valuable guidance and kind supervision given to us throughout the course which shaped the present work as its show.

REFERENCES

[1]. Arushi Tetarbe , Tanupriya Choudhury , Teoh Teik Toe , Seema Rawat," Oral cancer detection using data mining tool", 2017 3rd International Conference on Applied and Theoretical Computing and Communication Technology (iCATccT)

[2]. Bourass Youssef, Zouaki Hamid, Bahri Abdelkhalak, "Computer-aided diagnostics of facial and oral cancer", *IEEE Transactions,* 2015 Third World Conference on Complex Systems (WCCS)

[3]. Konstantina Kourou, Themis P. Exarchos, Konstantinos P. Exarchos, Michalis V. Karamouzis, Dimitrios I. Fotiadis, "Computational and Structural Biotechnology Journal", Machine learning applications in cancer prognosis and prediction, Volume 13, 2015, Pages 8-17

[4]. Harikumar Rajaguru, Sunil Kumar Prabhakar, "Performance Comparison of Oral Cancer Classification with Gaussian Mixture Measures and Multi-Layer Perceptron" ,The 16th International Conference on Biomedical Engineering,30 June 2017, pp 123-129

[5]. K.P. Exarchos, Y. Goletsis, D.I. Fotiadis "Multiparametric decision support system for the prediction of oral cancer reoccurrence", IEEE Trans Inf Technol Biomed, 16 (2012), pp. 1127-1134

[6]. Pandia Rajan Jeyaraj, Edward Rajan Samuel Nadar, "Computer-assisted medical image classification for early diagnosis of oral cancer employing deep learning algorithm", Journal of Cancer Research and Clinical Oncology 145, pages829–837(2019)

[7]. R Rajkumar, Dr. M V Sudhamani and Head Department of ISE, RNSIT Bengaluru, Karnataka, India, "Crawler for Image Acquisition from World Wide Web", International Journal of Engineering Trends and Technology (IJETT) –Special Issue-May 2017ISSN: 2231-5381

[8]. Mingyuan Xin, Yong Wang, "Research on image classification model based on deep convolution neural network", EURASIP Journal on Image and Video Processing volume 2019, Article number: 40 (2019)

DOI 10.17148/IJARCCE.2020.9335